International Journal of Applied Science and Engineering

Lightweight model based on improved YOLOv7 tiny for potato leaf diseases detection

Li-Hua Li [1], **Radius Tanone** [1,2*]

[1] *Department of Information Management, Chaoyang University of Technology, Taichung City, Taiwan*
[2] *Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Indonesia*

ABSTRACT

Indonesia, a country heavily dependent on agriculture, continues to grow potatoes. However, the presence of plant diseases, manifested by the condition of the leaves, is a significant problem that requires attention. Agriculture offers extensive opportunities to explore computer vision applications, including tasks like object detection. In this paper, we present a method that increases the YOLOv7 tiny model's accuracy to assist farmers in identifying diseases in potato leaves. Our study employed multi-scale and MixUp augmentation techniques to process input images when training using the YOLOv7 tiny model. Based on our experiment, the model can be enhanced using multi-scale training instead of fixed-scale training. After implementing our proposed technique, the mAP metric significantly improved over the original model, achieving a range of 0.94325 to 0.96975 for fixed-scale training and a range of 0.9620 to 0.97525 for multi-scale training with the MixUp approach. In addition, we have developed the YOLOv7 tiny model, which aims to enable seamless use of mobile devices in real-time applications. To assess the current state of potato leaves on land in real-time, we convert the results of our extended model into a compact format called TF Lite. Future potato production can be improved by using these findings to help farmers combat leaf diseases.

Keywords: MixUp, Multi-scale, Potato leaf diseases, TF Lite, YOLOv7 tiny.

OPEN ACCESS

Received: January 22, 2024
Revised: February 2, 2024
Accepted: February 15, 2024

Corresponding Author:
Radius Tanone
radius.tanone@uksw.edu

Publisher:
Chaoyang University of Technology
ISSN: 1727-2394 (Print)
ISSN: 1727-7841 (Online)

1. INTRODUCTION

Agriculture is a critical factor that can affect any country's economic growth, including Indonesia (*The Digital Transformation of Agriculture in Indonesia*). Agriculture in Indonesia covers 70 million hectares, of which only 45 million are ready for use (*DIPERTAPA - Kebutuhan Lahan Untuk Pangan Capai 13,17 Juta Ha*). The available data indicates a consistent decline in the agricultural land area over time. On the other hand, this paradox poses a new challenge for an agrarian country like Indonesia to survive in its agricultural sector, which remains one of its pillars. In addition, other facts show that crops in Indonesia are very diverse. Crops commonly used as a mainstay of agriculture in Indonesia include rice (Mariyono, 2019), corn (Suriani et al., 2021), cassava (Sukara et al., 2020; Hasiholan et al., 2021), or vegetable crops (Mariyono, 2020; Mariyono et al., 2020) such as kale, spinach, lettuce, and others. Potato production is the goal of the Indonesian government to increase food demand and economic growth. However, diseases in potatoes identified by leaves are sometimes one of the main factors that trigger potato production failure (Yusianto et al., 2020; Taylor and Dawson, 2021; Saptana et al., 2022) , thus failing to achieve the target. Various factors, such as weather, soil conditions, and fertilizers, cause this disease. However, the challenge of recognizing diseases in potato leaves can be a problem farmers face. In more detail, the condition of the leaves affected by disease makes the shape or color and even the type of disease challenging to distinguish from one another. In this case, technological assistance such as object detection could be one solution to help farmers more effectively recognize types of disease on potato leaves.

Furthermore, more computational technological interventions are needed in the agricultural sector to help farmers identify plant disease that can disrupt crops. Deep learning models are increasingly used in image or video processing (Laghari et al., 2023) to produce results that help people solve real-world problems. Identification of objects in images (Laghari et al., 2023), especially in agricultural fields for plant disease detection (Eldeen et al., 2021), has become popular in computer vision, equal to image classification (Saeed et al., 2023) or object detection (Zhang et al., 2020; Ouhami et al., 2021). As in our previous research, the image classification task for potato leaf disease classification can be solved by implementing a deep learning model based on a transformer-based model such as Swin Transformer (Li and Tanone, 2023). However, future improvement must address potato farmers' pressing issues related to video processing, such as object detection using models like You Only Look Once (YOLO) framework. In addition, there are many mobile phone users (Ahmad et al., 2024; Prabhu N and Majhi, 2024), including farmers. In further development, integrating mobile and other devices in an architecture such as IoT for agriculture has become an exciting research topic. This raises the question, is it possible to detect disease on potato leaves based on object detection using a model like YOLOv7 tiny? Of course, exploring how this tiny model is further processed into a TF Lite-based model, which can be used to build mobile applications based on iOS, android, or other mobile operating systems, would be more attractive. The impact of developing this idea can help farmers detect diseases on potato leaves based on mobile in the future.

Based on the background, we are interested in further investigating how the problem of disease identification in potato leaves can be addressed using a model such as YOLOv7 tiny. Our motivation is to build the lightweight model by improving the YOLOv7 tiny model to assist potato farmers in identifying disease detection on potato leaves. In line with this motivation, we use multi-scale techniques and adjust advanced data augmentation like Mosaic and MixUp to improve the accuracy of the YOLOv7 tiny model. After that, we developed this model into a mobile-based model for real-time applications. In addition, we have a limitation in this research: we only prepared the model on the TF Lite version but did not test it on mobile devices as we prepared for our future research.

In detail, our contribution to this research is presented as follows:

1. We examined the YOLOv7 tiny model and made improvements to this model using a multi-scale resolution training approach. In addition, we improve the model by adjusting data augmentation techniques like Mosaic and MixUp to improve the model performance.
2. Our study compared the tiny model trained with a fixed-scale and a multi-scale input image in YOLOv7 tiny. Using this approach, we can examine how the model's performance impacts the detection of potato leaf conditions.
3. To detect potato leaf conditions in real-time applications, we also created a lightweight model by converting it to a TF Lite model, and we examined the model. In the future, this model can be implemented in mobile applications to detect potato leaf diseases.

Furthermore, this paper will discuss related works followed by methods for improved YOLOv7 tiny model to detect potato leaf diseases in section 2. The results and discussion of the experiments performed are presented in section 3. The rest are the conclusions and references of this paper.

2. RELATED WORKS

Since YOLO was first launched in 2015 by Redmon et al. (2015), much research has been done on YOLO. In fact, until recently, the latest version of YOLO was YOLOv8, developed by Ultralytics (YOLOv8 - Ultralytics | Revolutionizing the World of Vision AI). However, in several previous studies, YOLOv7 still dominates in object detection. As reported by Pham et al. (2022), which proposes to collect and label road damage data using Google street view and use YOLOv7 together with coordinate attention and related accuracy fine-tuning techniques such as label smoothing and ensemble method to train deep learning models for automatic road damage detection and classification. This approach produces accurate results in testing 74.1% of the dataset used. YOLOv7's excellent performance makes it possible to implement object detection in agricultural areas. One of the research studies was conducted by Gallo et al. (2023). Deep weed object detection was accomplished by putting the most recent YOLOv7 to the test on both the chicory plant (CP) and Lincoln beet (LB) datasets, for which a previous version of YOLO was used to map weeds and crops. Furthermore, the CP dataset was trained using the YOLO version and others with mean average precision (mAP)@0.5 scores, recall, and precision, which are low. However, after using YOLOv7 on the LB dataset, the values increased the mAP@0.5 scores from 51% to 61%, 68% to 74%, and 35% to 48% for the total mAP, mAP for weeds, and mAP for sugar beets, respectively.

Furthermore, some researchers on object detection in agriculture also use YOLOv7, for example, Wu et al. (2022), which detects camellia oleifera fruit in complex scenes by using YOLOv7 and data augmentation. In their experiment, the data augmentation YOLOv7 (DA-YOLOv7) model was created by combining the YOLOv7 network with various data augmentation methods. With mAP, precision, recall, f1 score, and average detection time of 96.03%, 94.76%, 95.54%, 95.15%, and 0.025 s per image, the DA-YOLOv7 model had the best detection performance and a strong generalization ability in complex scenes. This is undoubtedly another finding, given YOLOv7's good performance detecting objects in agricultural areas. The same thing was done by Gumma and Yuan (2023), who compared YOLOv7 and YOLOv4 regarding image annotation quality for apple flower bud classification. In his research, YOLOv7 outperformed YOLOv4 for all growth stages and all training image annotation quality levels on the same test dataset. YOLOv7 achieved a 0.80 mAP with 100% training image annotation quality but only a 0.63 mAP with only 5% training image annotation quality. Depending on the apple flower bud growth stage and training image annotation quality, YOLOv7 improved YOLOv4 APs by 1.52% to 166.48% and mAPs by 3.43% to 53.45%.

Moreover, Siddique et al. (2023) conducted research using the YOLOv7 tiny model, in which the YOLOv7 tiny approach was used in the Jetson nano edge device to detect Bangla sign language in real-time. As a result of the proposed system, Bangladeshi people with hearing impairments will be able to easily communicate effectively, simply, and cost-effectively. Various deep learning models and architectures have been developed concerning object detection. One of the best-known is YOLOv7 (Wang et al., 2022), where this model can run computational processes that detect objects in images or videos faster and more accurately. The previous version, YOLOv5, was also published by Dai et al. (2022) and used to detect potato leaf diseases. They proposed a hybrid YOLO v5 model with data augmentation and activation of the compression mechanism to identify potato disease.

In contrast to previous research, our research focuses on building mobile-based models using YOLOv7 tiny. Based on our proposed method, we train the model using multi-scale techniques and adjust advanced data augmentation like Mosaic and MixUp to improve the accuracy of the YOLO tiny model. These results are certainly a recommended solution for potato farmers to be more efficient in recognizing types of disease on potato leaves.

3. MATERIAL AND METHODS

3.1 Deep Learning and Object Detection

Besides machine learning in agriculture (Maya Gopal and Bhargavi, 2019; Liu, 2020; Ben Ayed and Hanana, 2021), deep learning (LeCun et al., 2015; Popkova, 2022) is one of the parts of AI that is the basis for solving computer vision problems. Deep learning, multi-layered computing in decision-making, is a trend that continues to evolve for the better. Types of deep learning such as supervised, unsupervised, and reinforcement learning (RL) are driving researchers to continue developing more optimal models. Object detection is a computer vision technique that locates objects in images or videos. Object detection algorithms typically use machine learning or deep learning to produce meaningful results. We can quickly recognize and locate objects of interest when we look at images or videos. Object detection aims to use a computer to replicate this intelligence. Object detection (Tong et al., 2020; Joseph et al., 2021; Zou et al., 2023) has many applications in computer vision, including detecting leaf conditions in agriculture. In the internet of things (IoT) age, agricultural sectors also keep up with developments by using devices such as drones, webcams, or smartphones to monitor crop development. Potato leaf detection with the YOLOv7 tiny model could be one of the tools available to farmers to detect the condition of potato leaves in real-time.

3.2 YOLOv7

In July 2022, YOLOv7 (Wang et al., 2022) was released, and it achieves state-of-the-art performance and is trained to detect the generic 80 classes in the MS COCO dataset for real-time object detection. The model has six variants, ranging from the YOLOv7 (the fastest, smallest, and least accurate) to the beefy YOLOv7-E6E. (slowest, largest, and most accurate). The differences between the model sizes are the image input resolution, the number of anchors, the number of parameters, and the number of layers. Moreover, basic YOLOv7 consists of the backbone, neck, and head in Fig. 1.

The E-elan based on elan (Zhang et al., 2022) is the foundation of YOLOv7. The E-elan architecture of YOLOv7 allows the model to learn more effectively while maintaining the original gradient route by employing the "expand, shuffle, and merge cardinality". Moreover, E-elan only modifies the architecture of the computational block, leaving the architecture of the transition layer untouched. The E-elan strategy uses group convolution to increase the number of computational blocks and the channel. It uses the same group parameter and channel multiplier for all computational blocks within a computational layer. The feature map computed by each computational block is then shuffled into g groups based on the set group parameter g and concatenated together. The number of channels in each group of feature maps will be the same as in the original architecture at this time. Finally, add g feature map groups to perform merge cardinality. E-elan can guide different groups of computational blocks to learn more diverse features while maintaining the original elan design architecture.

3.3 Flowchart Proposed Method

Our methodology for conducting this experiment consists of three phases: data, training, and inferences. Fig. 2 shows the specifics of our methodological process.

In Fig. 2, the first stage is preparing a dataset. Our experiment took the potato leaf dataset from the Roboflow open dataset (*Roboflow Universe: Open Source Computer Vision Community*). The data format has been adapted to YOLOv7 from the Roboflow website. Furthermore, we store the data we use in this experiment on GitHub (GitHub -Kakarads/YOLOv7-potato-leaf: Dataset YOLOv7-potato-leaf). After the dataset has been obtained, the next step is pre-processing to prepare the data in YOLOv7 format. We carefully checked the collected data and then divided it into two parts: training data and test data. Our split data ratio is 80 : 20 for training and testing. In addition, data annotation is also performed to meet the criteria for deep learning computational processes on YOLOv7. In this data phase, we make parameter adjustments to customize the dataset that will be processed later in the training process. In this phase of data expansion, we perform three types of expansion, namely Mosaic augmentation (Wei et al., 2020) (with a 5% probability of MixUp augmentation), which is used for the basic YOLOv7 tiny model (training 1). Next is the augmentation MixUp without Mosaic augmentation which will be tried to do training 2 in the next step. Finally, we use the augmentation mix data between Mosaic and MixUp augmentation to see the performance of the YOLOv7 tiny model. The augmentation dataset is trained in the training 1, 2, and 3 schemes, as shown in Fig. 2. The purpose of these different training settings is to compare the training results and determine which augmentation is most effective. The next phase is inference, which comes after training.

The purpose of the inference phase is to evaluate the trained model. The model we have successfully trained is used to test the object detection performance of the tiny model used in the form of images and videos. Based on Fig. 2, we performed the inference stage using the YOLOv7 format model, then converted to ONNX format and drew conclusions. The final step was to convert the model to TF Lite format to use a lightweight model. In the following subsections, we explain the details of the individual processes. In addition, the results of this methodology can be viewed in the results section.

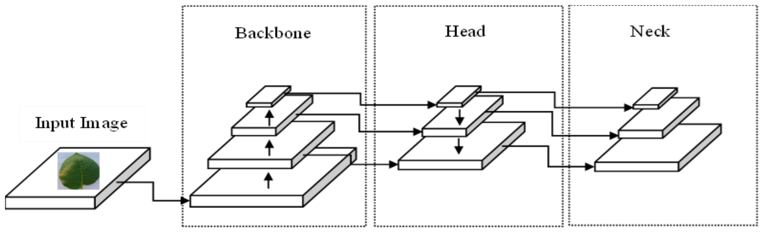


Fig. 1. YOLOv7 basic architecture

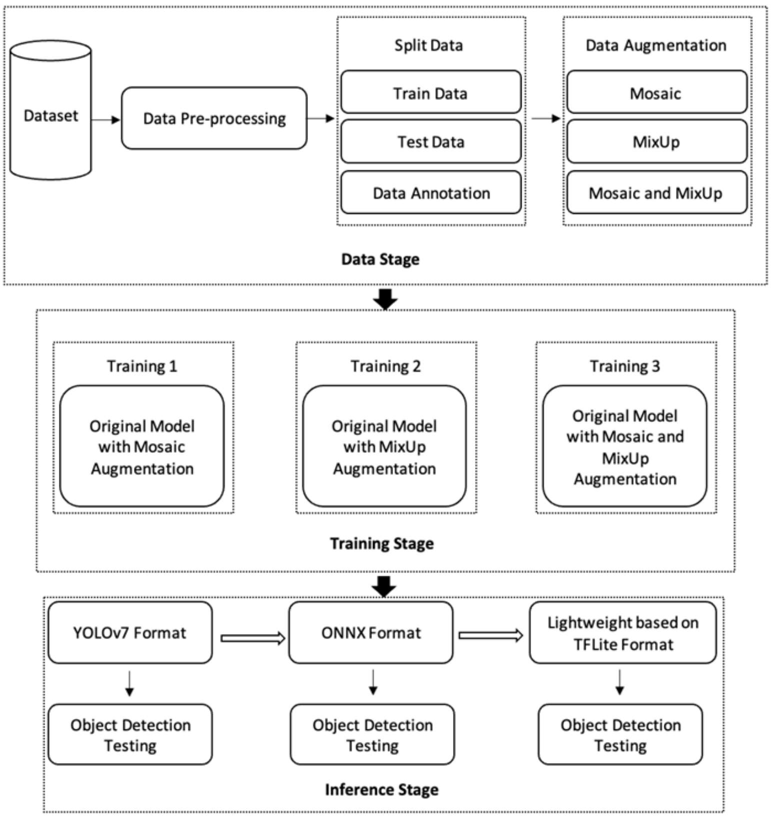


Fig. 2. Proposed methods

3.4 Dataset and Experimental Setup

In conducting this experiment, we used an open dataset from Roboflow. The description of the dataset furthermore, we convert the results of our improved model into TF Lite format for lightweight that can be used in real-time on land to detect the condition of potato leaves in real-time, which consists of three classes, which can be seen in Table 1. The dataset was split into training and validation, which consisted of 4133 and 1034 images, respectively. Also, we use the input image with a resolution of 640 × 640, which will later be resized to the resolution in the training with multi-scale.

In addition, we use environmental settings related to the hardware and software used to train the model to process this experiment. We typically use an Intel 11th generation CPU with an i9 11900K processor, NVIDIA GeForce RTX 3080 GPU, and 32GB of RAM.

Furthermore, we set the parameters for data augmentation to see the performance of the training model. More details can be seen in Table 2. Table 2 illustrates the augmentation data settings where the original model uses a 100% probability for Mosaic augmentation and 5% MixUp. In this original model, the dominant augmentation uses Mosaic. This performance needs to be improved by using MixUp augmentation. So, we also use a combination (100% probability) of Mosaic and MixUp and 100% MixUp in conducting training.

Table 1. Potato leaf dataset

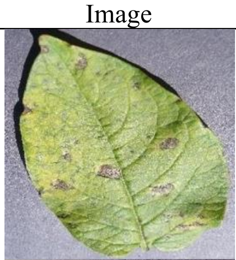
No.	Class	Image	Resolution
1	Early blight		640 × 640
2	Late blight		640 × 640
3	Healthy		640 × 640

Table 2. Parameters setting for data augmentation based on Mosaic and MixUp

Model	Training	Mosaic probability	MixUp probability
Original tiny model	Training 1	100%	5%
Custom tiny model	Training 2	-	100%
Custom tiny model	Training 3	100%	100%

3.5 Model and Parameters

The model used in this experiment is the YOLOv7 tiny model to identify diseases on potato leaves. The description of the model, along with the parameters when we do the training, is the default of the original YOLOv7 tiny model. In addition, while conducting this experiment, we also changed some parameter values to improve model performance in detecting the condition of potato leaf. We will go into more detail on data augmentation in the following subsection.

3.6 Fixed-scale and Multi-scale

When training, we combine fixed-scale and multi-scale strategies to achieve more accurate object detection results with the YOLOv7 tiny model. Fixed-scale means that the input image used for training has only one size, 640 × 640 pixels. To improve accuracy in recognizing potato leaves, we use variations in the size of the input image ranging from 640 to 1024 pixels in multi-scale. Details on the use of fixed and multi-scale during training are described in detail in Table 4. In detail, the number of epochs on the multi-scale is 300 because the training time needs to be longer to get maximum results. The results of the longer training effect can be viewed in the results section.

3.7 MixUp Augmentation

Data augmentation was introduced in 1998 (Simard et al., 1998) and formalized by some researchers (Chapelle et al., 2000), another data augmentation technique named MixUp by (Zhang et al., 2017). In processing the training data, the augmentation technique used when training the model is MixUp, not Mosaic. MixUp was chosen because it linearly interpolates input examples and the corresponding labels. Thus, Mixup interpolates linearly in the input space and similarly in the associated target space. This improves model robustness to corrupt labels, avoids overfitting because virtual labels are difficult to memorize, and increases generalization. In processing potato leaf as an input image, the part of the leaf that indicates disease is usually on the tip side. If the information from the input image is mixed randomly, this will undoubtedly be the robustness of the process in the YOLOv7 tiny model.

In order to improve the performance of the tiny model, MixUp is being used as the primary data source for the following data augmentation tasks: to the exact dimensions and resize the images. Take a sample from the Beta distribution to get the value, multiply all the values in image 1 by image 2 by 1, and so forth, and combine the annotations to create the final annotations for the image after adding the two images. Equations (1) and (2) of MixUp used to perform data augmentation.

$$\tilde{x} = \lambda x_i + (1 - \lambda)\ x_j \tag{1}$$

$$\tilde{y} = \lambda y_j + (1 - \lambda)\, y_j \tag{2}$$

Where x_i, x_j are raw input vectors and y_i, y_j are one-hot label encodings. Furthermore, note that the lambda values are values with the [0, 1] range and are sampled from the Beta distribution. To see the augmentation results using MixUp on the potato leaf dataset can be seen in Fig. 3.

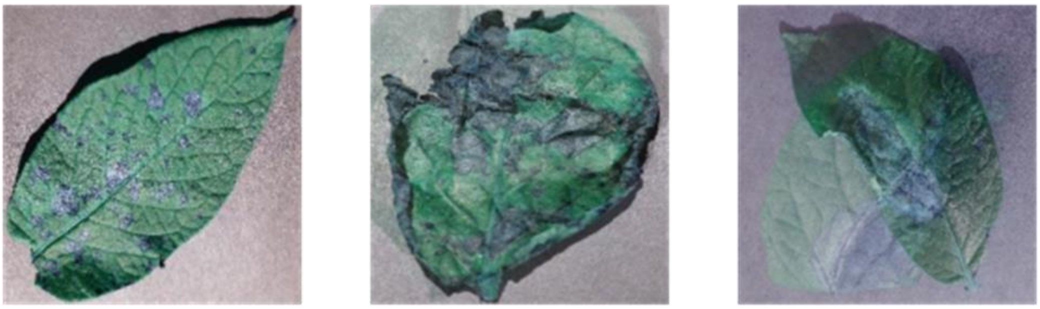


Fig. 3. MixUp image

3.8 TensorFlow Lite

According to the TensorFlow website (*TensorFlow Lite | ML for Mobile and Edge Devices*), TensorFlow Lite is a mobile library for deploying models on mobile, microcontrollers, and other edge devices. In this experiment, in addition to training the model to achieve high accuracy, inference must be made. We decided to turn the YOLOv7 tiny model into TF Lite in our research so that it can be built on top of, for example, Android-based real-time applications. When preparing the TF Lite file, we first convert the tiny model of YOLO to ONNX format and then convert it to TF Lite format. Another goal is to enable farmers to use smartphones to detect the condition of potato leaves using the YOLOv7 tiny model.

3.9 Metrics evaluation

The mAP metric was used to evaluate this YOLOv7-based model. In addition, the calculation of mAP requires intersection over union (IOU), precision, recall, precision recall curve, and average precision (AP). Object detection models predict the bounding box and category of objects in an image. IOU determines if the bounding box was correctly predicted. The IOU indicates how much-bounding boxes overlap. This overlap ratio between the areas of two bounding boxes becomes 1.0 in case of an exact match and 0.0 when there is no overlap. IOU formula can be seen in Equation (3).

$$IOU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{3}$$

Precision is a model's ability to identify only the relevant objects. A model that does not produce false positives has an accuracy of 1.0. However, the value is 1.0 even if there are undetected or unrecognized bounding boxes that should be detected. The precision formula can be seen in Equation (4).

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{All\ detections} \quad (4)$$

The ability of a model to find all ground truth bounding boxes on potato leaf images is called recall. A model with a recall of 1.0 produces no false negatives (undetected bounding boxes that should be detected). Even if an "over detection" and the incorrect bounding box are detected, the recall will remain at 1.0. Recall formula can be seen in Equation (5).

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{All\ ground\ truths} \quad (5)$$

The precision-recall curve is a graph that shows precision on the vertical axis and recall on the horizontal axis, while the mAP is calculated by setting the confidence threshold. The mAP is calculated by calculating each class's average precision (AP) and then averaging it across multiple classes. The mAP considers both false positives (FP) and false negatives (FN) and incorporates the trade-off between precision and recall (FN). Because of this property, mAP is a good metric for most detection applications. mAP formula can be seen in Equation (6).

$$mAP = \frac{1}{n}\sum_{i=1}^{n} APi \quad (6)$$

4. RESULTS AND DISCUSSION

4.1 Results

We conduct the training using two techniques, namely fixed-scale, and multi-scale. For multi-scale, we use more iterations to get maximum results from our improved model. To see the results of the improvements to the model we trained, we used the recall, precision, and mAP metrics to see the model's effectiveness in detecting conditions on potato leaves.

To detect the condition of potato leaves, we use a tiny model that has been trained. The results of comparing the trained original tiny and tiny models are shown in Fig. 4 (a, b, and c). The detection of the potato leaf image is shown in Fig. 4(a), where the original model was not very detailed in the detection because only one class was displayed. Even though the potato leaves in this image are in various states. There is an improvement from the model in performing detection, where more than one class has been successfully detected in Fig. 4(b). Furthermore, with multi-scale, the model could detect more potato leaf conditions in Fig. 4(c) compared to the others.

Overall, the images used to detect potato leaf conditions show that the improved model we built significantly impacted the detection of potato leaf conditions. The original YOLOv7 model could not detect the condition of the leaves in detail, while the improved model could detect the condition of the potato leaves in more detail.

(a) (b) (c)

Fig. 4. (a) original model, (b) fixed-scale training, (c) multi-scale training

(a) (b) (c)

Fig. 5. (a) original model, (b) fixed-scale model, (c) multi-scale model

Following that, we created a test to detect the condition of the potato leaves using videos we created ourselves. The results of the inference on the potato leaf video are shown in Fig. 5. For the YOLO7 tiny model, which we improved by using the MixUp augmentation and multi-scale training, the video detection accuracy is better than the usual tiny model. Fig. 5 shows the improved inference results of the tiny model. As seen in the three images (a, b, c), the detection in the exact second in the video differs from the detection by the model. The normal (a) tiny model cannot detect leaf potatoes, then the tiny model trained on a fixed-scale (b) image detects leaf blight states, but when the tiny model trained on a multi-scale (c) image, the detection is carried out more details. This increase occurs because MixUp augmentation and training use multi-scale images

Moreover, the results of comparing the YOLOv7 tiny model can be seen in Table 3. For more metrics performance, see Figs. 6–9. The original tiny model with the original augmentation data parameter settings resulted in a mAP of 0.94325 after 100 iterations during training. In this iteration with the original model, the input image is trained using fixed size, 640 × 640. In 300 iterations, where the input image was made multi-scale with between 320 × 320 and 1024 × 1024, the model produced an accuracy of 0.96200. To increase the model's accuracy, we tried to make a hybrid augmentation where, besides Mosaic, we added MixUp augmentation to the input image to be trained. As a result, these two augmentation data combinations did not work well in the potato leaf dataset, causing the model performance to decrease to 0.90975 and 0.90625 for 100 and 300 iterations, respectively. Seeing this change, according to the methodology we proposed in this research, we change the augmentation only by using MixUp for 100 and 300 iterations. We also use a training strategy with augmentation data using a fixed-scale and multi-scale for image input. Furthermore, this strategy can increase the performance of the tiny model to 0.96975 and 0.97525. This improved accuracy is due to the MixUp augmentation, which we will discuss in the discussion section.

Table 3. Comparison of mAP of different improved models

Model	Method	Epochs	Recall	Precision	mAP@@.5 : .95
YOLOv7 tiny	Original with fixed-scale	100	0.9877	0.9987	0.94325
YOLOv7 tiny	Original with multi-scale	300	0.9967	0.9987	0.96200
YOLOv7 tiny	Augmented (MixUp)	100	0.9937	0.9903	0.96975
YOLOv7 tiny	Augmented (MixUp) multi-scale	300	0.9960	0.9930	0.97525

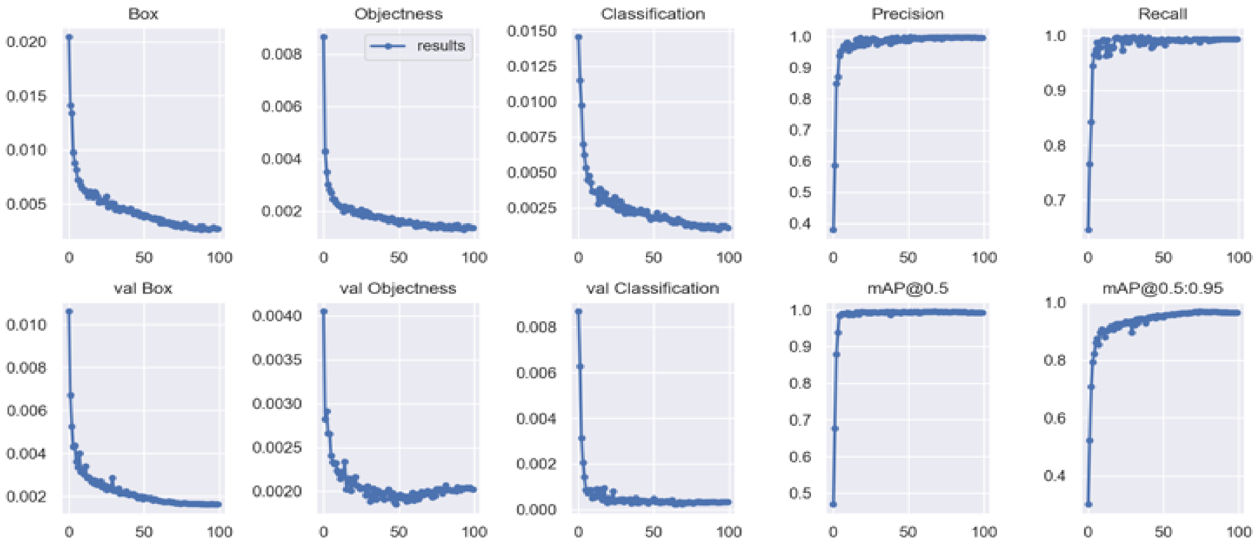


Fig. 6. Metrics performance for fixed-scale training with MixUp augmentation

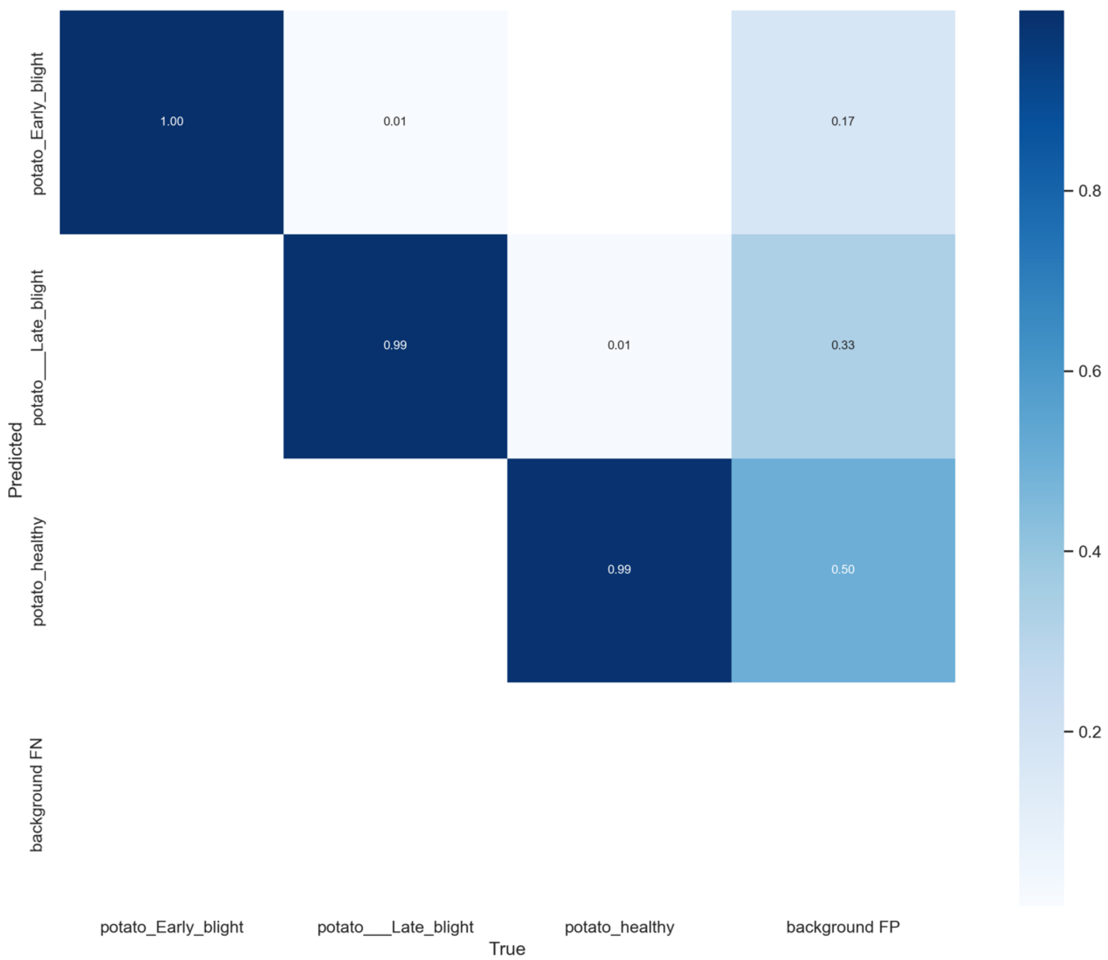


Fig. 7. Confusion matrix for fixed-scale training with MixUp augmentation

International Journal of Applied Science and Engineering

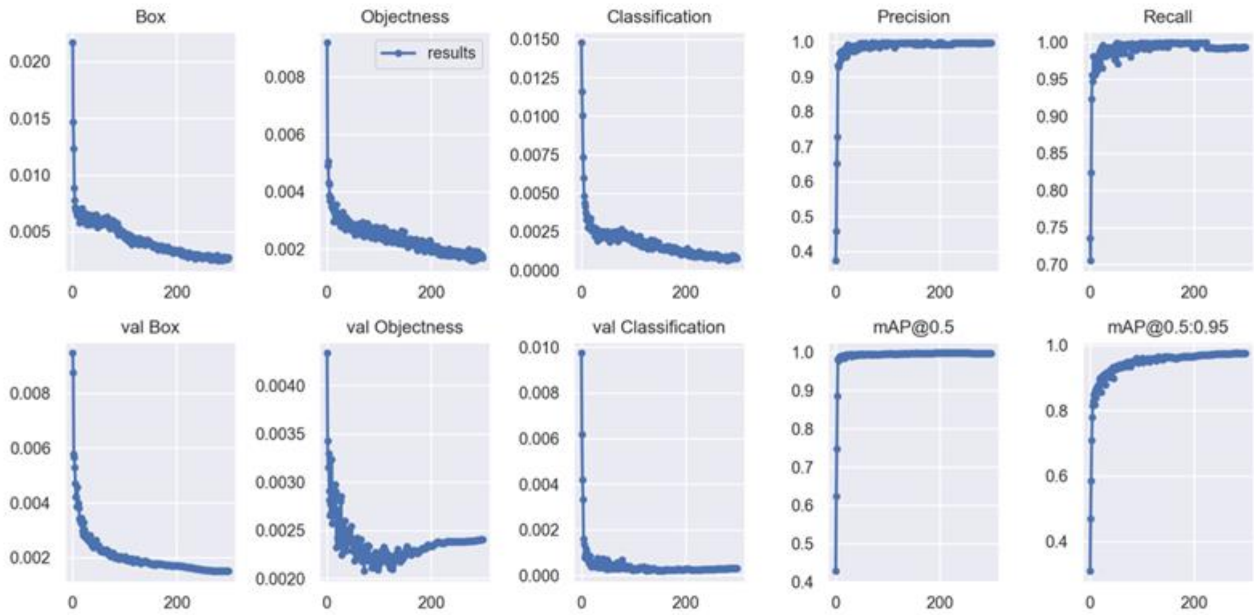


Fig. 8. Metrics performance for multi-scale training with MixUp augmentation

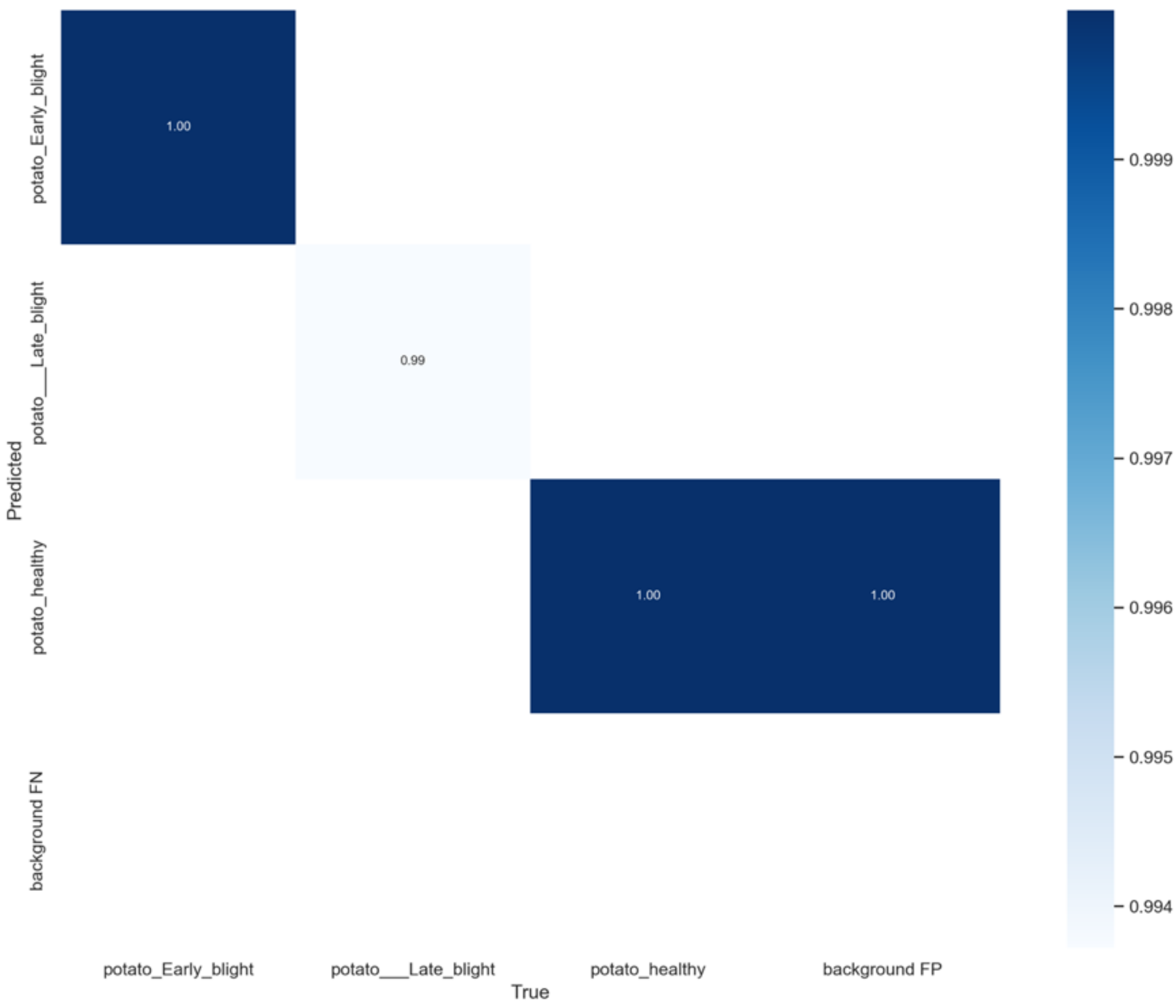
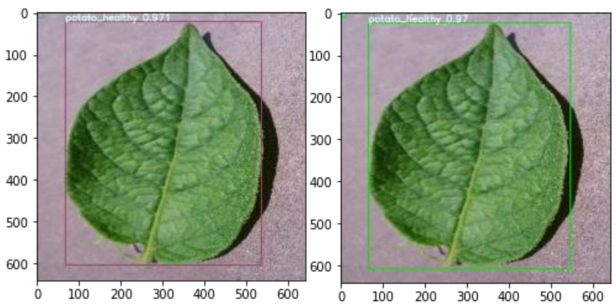


Fig. 9. Confusion matrix for multi-scale training with MixUp augmentation

After improvising the tiny model's performance, the next step is to make inferences using a TF Lite-based model. Fig. 10 shows the results of the inference model in TF Lite format. To get a TF Lite-based model, the first step is to convert the YOLOv7-based model to an ONNX-based model. Next, the ONNX format is changed to TF Lite. The TF Lite results in Fig. 10 perfectly predict healthy potato leaves. Furthermore, to compare the performance of the TF Lite-based model where the model is trained using a multi-scale input image, it can be seen in Fig. 10 (a) and (b).

In conclusion, converting a YOLOv7 tiny model to TF Lite brings several advantages for deploying object detection capabilities on resource-constrained devices. The process involves leveraging TF Lite's optimization techniques, such as quantization and model compression, to reduce the model size and ensure efficient inference. This enables real-time object detection on mobile devices while maintaining a low memory and computational footprint. By utilizing TF Lite, developers can achieve fast and reliable on-device inference, allowing for applications like real-time image processing and object recognition without a constant internet connection. The platform independence of TF Lite facilitates deployment across various devices and operating systems, contributing to the solution's versatility.

Additionally, the TF Lite ecosystem offers tools and support for custom operators, enabling fine-tuning and optimization specific to the YOLOv7 tiny model or other object detection architectures. The ongoing community contributions to TF Lite ensure developers have access to the latest advancements, making TF Lite a valuable framework for implementing efficient and accurate mobile object detection solutions, especially in agriculture for potato leaf disease detection.



(a) (b)

Fig. 10. (a) Fixed-scale inference, (b) multi-scale inference

4.2 Discussion

As a result of MixUp augmentation, the YOLOv7 tiny model proved more robust. Three conditions of potato leaves, namely early blight, healthy, and late blight, were mixed up randomly using MixUp augmentation. In more detail, MixUp generates combinations of random image pair weights from training data on potato leaf images. As illustrated in Fig. 3, the random combination allows the neural network to increase the generalization of the condition detection of potato leaves. This is because the randomness in the image pair allows the joining of each class in the image, such as early blight and healthy, healthy, and late blight, or early blight and late blight.

Furthermore, the robustness of the model shows an increase in accuracy in making predictions by conducting training using multi-scale images. We conducted a study that found that with a longer training time, multi-scale on the trained image can improve the model's performance in performing detection. This is because the image size is made from the smallest (320 pixels) to the largest (1024 pixels) during training. We conducted a study that showed that, with varying sizes, the model will learn better in identifying familiar potato leaf objects.

In our study, we successfully used MixUp augmentation rather than Mosaic augmentation or combined these two augmentations because the results were not better in comparison. The results using the Mosaic and MixUp augmentation (training using an original tiny model) can be seen in Table 4. The results obtained from combining these two augmentations make the mAP lower than the basic tiny model.

To conclude the discussion, seeing the performance improvement of the YOLOv7 tiny model in our experiments, the results of the model's conversion can be further referenced using mobile platform-based devices. The purpose of this inference is to be applied in real-time by farmers in agriculture. Furthermore, when it is used, the results of detecting the condition of potato leaves can be known quickly so that decisions can be made to help the production process and better harvests.

In carrying out this experiment, we also carried out comparisons with previous state-of-the-art research. In comparison, the techniques used by previous researchers are undoubtedly different, and the datasets used are also different. However, to see how object detection research is regarding potato leaf detection, we present it in Table 5.

Table 4. Result using Mosaic and MixUp augmentation

Model	Method	Epochs	Recall	Precision	mAP@@.5 : .95
YOLOv7 tiny	Augmented (Mosaic-MixUp) - fixed-scale	100	0.9790	0.9760	0.90975
YOLOv7 tiny	Augmented (Mosaic-MixUp) - multi scale	100	0.9703	0.9477	0.90625

Table 5. Comparison with state of the art research on potato leaf detection

Works	Dataset	Technique	Average precision
Oishi et al., 2021	Potato leaf dataset	Portable video and deep learning models	0.9090
Mohandas et al., 2021	Tomato, mango strawberry, beans and potato leaves	YOLOv4-tiny	0.6331
Chairma Lakshmi et al., 2023	Plant diseases dataset (potato leaf)	YOLOv3	0.9390
This study	Potato leaf dataset	YOLOv7 tiny original	0.9753

Table 5 illustrates that our experiment is better at producing average precision. However, when comparing the result with other research, with all due respect, it must be underlined that in carrying out object detection, the techniques and datasets used in each experiment are different according to needs. In addition, to produce good performance in identifying diseases on potato leaves, it is necessary to develop other techniques to produce better model performance. The YOLOv7 tiny model is highly recommended for producing object classification performance with small resources. This is a recommendation in the future so that it can be implemented in the agricultural sector to overcome existing problems.

5. CONCLUSION

In this experiment, we improved the YOLOv7 tiny model to detect conditions in potato leaves. According to our findings, the tiny model in YOLOv7 can be improved for the potato leaf dataset if the training uses only one data augmentation, MixUp. When using the original tiny model with the Mosaic augmentation, the model's accuracy and inference do not improve as expected. Another strategy we employ, combining Mosaic and MixUp data augmentation, causes model accuracy and inference to fail. Another finding from our research is that multi-scale training with more epochs can help the model perform well in accuracy and inference. In addition, the focus of our experiments is on how to make the tiny model on YOLOv7 work properly on mobile devices. In this experiment, we successfully implement the TF Lite format for mobile devices, which is extremely useful for developers creating applications to detect potato leaf conditions. In terms of future development, the results of our improvement model can be used in the field by applying it to agricultural land using more embedded systems or drones.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

DECLARATIONS

Conflicts of Interests The authors declare that they have no conflict of interest.

REFERENCES

Ahmad, B., Sarkar, M.A.R., Khanom, F., Lucky, R.Y., Sarker, M.R., Rabbani, M.G., Ray, S.R.R., Rahman, M.N., Sarker, M.N.I. 2024. Experience of farmers using mobile phone for farming information flow in Boro rice production: A case of Eastern Gangetic Plain. Social Sciences & Humanities Open, 9, 100811.

Ben Ayed, R., Hanana, M. 2021. Artificial intelligence to improve the food and agriculture sector. Journal of Food Quality, 2021, 5584754.

Chairma Lakshmi, K.R., Praveena, B., Sahaana, G., Nithya Jenev, J., Gnanasekaran, T., Hashim, M. 2023. YOLO for detecting plant diseases. Proceedings of the 3rd International Conference on Artificial Intelligence and Smart Energy, ICAIS 2023, 1029–1034. https://doi.org/10.1109/ICAIS56108.2023.10073875

Chapelle, O., Weston, J., Bottou, L., Vapnik, V. 2000. Vicinal risk minimization. Advances in Neural Information Processing Systems. 13.

Dai, G., Hu, L., Fan, J. 2022. DA-ActNN-YOLOV5: Hybrid YOLO v5 model with data augmentation and activation of compression mechanism for potato disease identification. Computational Intelligence and Neuroscience, 2022, 6114061.

DIPERTAPA - Kebutuhan Lahan Untuk Pangan Capai 13,17 Juta Ha. Retrieved April 6, 2023, from https://pertanian.kulonprogokab.go.id/detil/10/kebutuhan-lahan-untuk-pangan-capai-1317-juta-ha

Eldeen, N., Khalifa, M., Hamed, M., Taha, N., Abou El-Maged, L.M., Hassanien, A.E. 2021. Artificial intelligence in potato leaf disease classification: A deep learning approach. Studies in Big Data. 63–79.

Gallo, I., Rehman, A.U., Dehkordi, R.H., Landro, N., La Grassa, R., Boschetti, M. 2023. Deep object detection of crop weeds: Performance of YOLOv7 on a real case dataset from UAV Images. Remote Sensing, 15(2), 539.

GitHub - kakarads/YOLOv7-Potato-Leaf: DATASET YOLOv7-Potato-Leaf. Retrieved February 17, 2024, from https://github.com/kakarads/YOLOv7-Potato-Leaf

Gumma, K., Yuan, W. 2023. Accuracy comparison of YOLOv7 and YOLOv4 regarding image annotation quality for apple flower bud classification. AgriEngineering, 5(1), 413–424.

Hasiholan, R.P., Basri Tarmizi, H., Rahmanta, R., Purwoko, A. 2021. Measurement instruments for superior product development: A case study of deli serdang cassava in Indonesia. The Journal of Asian Finance, Economics and Business, 8(5), 1139–1145.

Joseph, K.J., Khan, S., Khan, F.S., Balasubramanian, V.N. 2021. Towards open world object detection. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5826–5836.

Laghari, A.A., Shahid, S., Yadav, R., Karim, S., Khan, A., Li, H., Shoulin, Y. 2023. The state of art and review on video streaming. Journal of High Speed Networks, 29(3), 211–236.

Laghari, A.A., Sun, Y., Alhussein, M., Aurangzeb, K., Anwar, M.S., Rashid, M. 2023. Deep residual-dense network based on bidirectional recurrent neural network for atrial fibrillation detection. Scientific Reports, 13(1), 1–12.

LeCun, Y., Bengio, Y., Hinton, G. 2015. Deep learning. Nature, 521(7553), 436–444.

Li, L.H., Tanone, R. 2023. Disease identification in potato leaves using swin transformer. Proceedings of the 2023 17th International Conference on Ubiquitous Information Management and Communication, IMCOM 2023.

Liu, S.Y. 2020. Artificial intelligence (AI) in agriculture. IT Professional, 22(3), 14–15.

Mariyono, J. 2019. Farmer training to simultaneously increase productivity of soybean and rice in Indonesia. International Journal of Productivity and Performance Management, 68(6), 1120–1140.

Mariyono, J. 2020. Motivating factors of farmers to engage vegetable-based agribusiness in East Java and Bali, Indonesia. Jurnal Ekonomi Pembangunan: Kajian Masalah Ekonomi Dan Pembangunan, 20(2), 163–175.

Mariyono, J., Abdurrachman, H., Suswati, E., Susilawati, A.D., Sujarwo, M., Waskito, J., Suwandi, Zainudin, A. 2020. Rural modernisation through intensive vegetable farming agribusiness in Indonesia. Rural Society, 29(2), 116–133.

Maya Gopal, P.S., Bhargavi, R. 2019. Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms. Applied Artificial Intelligence. 33(7), 621–642.

Mohandas, A., Anjali, M.S., Rahul Varma, U. 2021. Real-time detection and identification of plant leaf diseases using YOLOv4-tiny. 2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021.

Oishi, Y., Habaragamuwa, H., Zhang, Y., Sugiura, R., Asano, K., Akai, K., Shibata, H., Fujimoto, T. 2021. Automated abnormal potato plant detection system using deep learning models and portable video cameras. International Journal of Applied Earth Observation and Geoinformation, 104, 102509.

Ouhami, M., Hafiane, A., Es-Saady, Y., El Hajji, M., Canals, R. 2021. Computer vision, IoT and data fusion for crop disease detection using machine learning: A survey and ongoing research. Remote Sensing, 13(13), 2486.

Pham, V., Nguyen, D., Donan, C. 2022. Road damage detection and classification with YOLOv7. 2022 IEEE International Conference on Big Data (Big Data), 6416–6423.

Popkova, E.G. 2022. Vertical farms based on hydroponics, deep learning, and AI as smart innovation in agriculture. Smart Innovation, Systems and Technologies, 264, 257–262.

Prabhu N.S., Majhi, R. 2024. An empirical investigation to understand mobile phone users' behavioural intention to give their end-of-life mobile phones for formal recycling. Waste Management, 177, 34–45.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A. 2015. You only look once: Unified, real-time object detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December, 779–788.

Roboflow Universe: Open Source Computer Vision Community. Retrieved August 17, 2023, from https://universe.roboflow.com/

Saeed, U., Kumar, K., Khuhro, M.A., Laghari, A.A., Shaikh, A.A., Rai, A. 2023. DeepLeukNet—A CNN based microscopy adaptation model for acute lymphoblastic leukemia classification. Multimedia Tools and Applications, 1–25.

Saptana, Sayekti, A.L., Perwita, A.D., Sayaka, B., Gunawan, E., Sukmaya, S.G., Hayati, N.Q., Yusuf, Sumaryanto, Yufdy, M.P., Mardianto, S., Pitaloka, A.D. 2022. Analysis of competitive and comparative advantages of potato production in Indonesia. PLOS ONE, 17(2), e0263633.

Siddique, S., Islam, S., Neon, E.E., Sabbir, T., Naheen, I.T., Khan, R. 2023. Deep learning-based bangla sign language detection with an edge device. Intelligent Systems with Applications, 18, 200224.

Simard, P.Y., LeCun, Y.A., Denker, J.S., Victorri, B. 1998. Transformation invariance in pattern recognition- tangent distance and tangent propagation, Neural Networks: Tricks of the Trade (pp. 239–274). Springer Berlin Heidelberg.

Sukara, E., Hartati, S., Ragamustari, S.K. 2020. State of the art of Indonesian agriculture and the introduction of innovation for added value of cassava. Plant Biotechnology Reports, 14(2), 207–212.

Suriani, Patandjengi, B., Junaid, M., Muis, A. 2021. The presence of bacterial stalk rot disease on corn in Indonesia: A review. IOP Conference Series: Earth and Environmental Science, 911(1), 012058.

Taylor, A.S., Dawson, P. 2021. Major constraints to potato production in Indonesia: A review. American Journal of Potato Research, 98(3), 171–186.

TensorFlow Lite | ML for Mobile and Edge Devices. Retrieved April 6, 2023, from https://www.tensorflow.org/lite

The digital transformation of agriculture in Indonesia.

Retrieved July 5, 2022, from https://www.brookings.edu/blog/future-development/2022/01/21/the-digital-transformation-of-agriculture-in-indonesia/

Tong, K., Wu, Y., Zhou, F. 2020. Recent advances in small object detection based on deep learning: A review. Image and Vision Computing, 97, 103910.

Wang, C.-Y., Bochkovskiy, A., Mark Liao, H.-Y. 2022. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Computer Science, 1–15.

Wei, Z., Duan, C., Song, X., Tian, Y., Wang, H. 2020. AMRNet: Chips augmentation in aerial images object detection. Computer Science, 1–5.

Wu, D., Jiang, S., Zhao, E., Liu, Y., Zhu, H., Wang, W., Wang, R., Wu, D., Jiang, S., Zhao, E., Liu, Y., Zhu, H., Wang, W., Wang, R. 2022. Detection of camellia oleifera fruit in complex scenes by using YOLOv7 and data augmentation. Applied Sciences, 12(22), 11318.

YOLOv8 - Ultralytics | Revolutionizing the World of Vision AI. Retrieved April 6, 2023, from https://ultralytics.com/yolov8

Yusianto, R., Marimin, Suprihatin, Hardjomidjojo, H. 2020. Spatial analysis for crop land suitability evaluation: A case study of potatoes cultivation in wonosobo, Indonesia. Proceedings - 2020 International Seminar on Application for Technology of Information and Communication: IT Challenges for Sustainability, Scalability, and Security in the Age of Digital Disruption, ISemantic 2020, 313–319.

Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D. 2017. Mixup: Beyond empirical risk minimization. 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings.

Zhang, X., Zeng, H., Guo, S., Zhang, L. 2022. Efficient long-range attention network for image super-resolution. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13677, 649–667.

Zhang, Y., Song, C., Zhang, D. 2020. Deep learning-based object detection improvement for tomato disease. IEEE Access, 8, 56607–56614.

Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J. 2023. Object detection in 20 years: A survey. Proceedings of the IEEE.